Improving bounds on invisible branching ratio of the Higgs with Deep-Learning


Physical Research Laboratory

XXIV DAE-BRNS High Energy Physics symposium
December 16, 2020

Vishal S. Ngairangbam
Motivation: Looking at VBF Higgs through a CNN

Invisible Higgs search at LHC

Data-representation: high-level and low-level features

Preprocessing

Network Performance

Result: Bounds on invisible branching ratio of Higgs

Back-up
Motivation: Looking at VBF Higgs through a CNN

Invisible Higgs search at LHC

Data-representation: high-level and low-level features

Preprocessing

Network Performance

Result: Bounds on invisible branching ratio of Higgs

Back-up
Vector Boson Fusion: A unique signature

- t-channel production of color-singlet particles via fusion of two vector-bosons
  - No central jet activity
  - Large rapidity gap between two jets
  - Large invariant mass of the two jet system
  - Decay products at the central region

- Higher order QCD always below 10% – very stable with scale uncertainty

- Very important for BSM searches of color singlet particles.

- Dominant production channel for heavy Higgs at hadron colliders

- **Central-jet veto**: viable to search for lighter Higgs masses
Vector Boson Fusion: A unique signature

▶ t-channel production of color-singlet particles via fusion of two vector-bosons
  ▶ No central jet activity
  ▶ Large rapidity gap between two jets
  ▶ Large invariant mass of the two jet system
  ▶ Decay products at the central region

**VBF production of** \( m_h = 125 \text{ GeV} \) **Higgs**

▶ Second highest cross-section after gluon-fusion

▶ Very clean channel for non-hadronic decay of the Higgs

▶ Most sensitive channel for searching invisible decay of Higgs (Important in many BSM scenario)
Vector Boson Fusion: A unique signature

- t-channel production of color-singlet particles via fusion of two vector-bosons
  - No central jet activity
  - Large rapidity gap between two jets
  - Large invariant mass of the two jet system
  - Decay products at the central region

Collider bounds on invisible branching ratio of Higgs much higher than in SM!!
Vector Boson Fusion: A unique signature

- t-channel production of color-singlet particles via fusion of two vector-bosons
  - No central jet activity
  - Large rapidity gap between two jets
  - Large invariant mass of the two jet system
  - Decay products at the central region

Collider bounds on invisible branching ratio of Higgs much higher than in SM!!

New techniques to reduce the upper limit: Deep learning??
CNNs and jet-images: why do they work?

- Efficiently distinguishes large radius QCD jets from decays of boosted heavy particles ($t, W^\pm/Z^0/h^0$)

- Works with data which have an underlying Euclidean-geometry

- Jet-substructure variables are mostly functions of the Euclidean distance:
  \[
  \Delta R_{ij} = \sqrt{\Delta \eta_{ij}^2 + \Delta \phi_{ij}^2}
  \]
  in the $(\eta, \phi)$ plane, for instance:

  \[
  \text{ECF}(2, \beta) = \sum_{i,j<i \in J} p_T^i p_T^j (\Delta R_{ij})^\beta
  \]
Salient underlying event structure in Vector-boson fusion (VBF): no color exchanged at LO

Can CNNs leverage information from the full calorimeter tower?
Salient underlying event structure in Vector-boson fusion (VBF): no color exchanged at LO

Can CNNs leverage information from the full calorimeter tower?

Turns out, we can!
Motivation: Looking at VBF Higgs through a CNN

Invisible Higgs search at LHC

Data-representation: high-level and low-level features

Preprocessing

Network Performance

Result: Bounds on invisible branching ratio of Higgs

Back-up
Search for Invisible decays of Higgs at LHC

- Higgs does not couple to $\nu$ in SM, couples to dark-matter in many BSM models

- Most recent ATLAS preliminary result$^a$ puts upper limit on $\text{B.R}(h \rightarrow \text{inv}) < 0.13$ at 95% confidence level with $\mathcal{L} = 140$ fb$^{-1}$.

- Reproduced the shape-analysis of CMS result$^b$ in our setting, for better comparison of increased sensitivity
  
  - deliberately weaken cuts in $|\Delta\eta_{jj}|$ and $m_{jj}$
  - Two signals: $S_{EW}$ (VBF) and $S_{QCD}$ (Gluon-fusion)

- We consider the following major backgrounds:
  
  - $Z_{QCD}$: $Z(\nu \bar{\nu}) + \text{jets}$
  - $W_{QCD}$: $W^{\pm}(l^{\pm}\nu) + \text{jets}$
  - $Z_{EW}$: VBF production of $Z(\nu \bar{\nu}) + 2 \text{jets}$
  - $W_{EW}$ : VBF production of $W^{\pm}(l^{\pm}\nu) + 2 \text{jets}$

---

$^a$ATLAS-CONF-2020-008

Pre-selection cuts

- **VBF Jet tag**: At least two jets with leading(sub-leading) jet $p_T > 80$ (40) GeV with $|\eta| < 4.7$. At least one of the jets to have $|\eta_{j_i}| < 3$.

  $\eta_{j_1} \eta_{j_2} < 0$, $|\Delta \phi_{jj}| < 1.5$, $|\Delta \eta_{jj}| > 1$, $m_{jj} > 200$ GeV

- **Lepton-veto**: No electron(muon) with $p_T > 10$ GeV in the central region, $|\eta| < 2.5(2.4)$.

- **Photon-veto**: No photon with $p_T > 15$ GeV in the central region, $|\eta| < 2.5$

- **$\tau$ and b-veto**: no tau-tagged jets in $|\eta| < 2.3$ with $p_T > 18$ GeV, and no b-tagged jets in $|\eta| < 2.5$ with $p_T > 20$ GeV.

- **Missing $E_T$(MET)**: $MET > 200$ GeV (250 GeV for CMS shape-analysis)

- **MET jet alignment**: $\min(\Delta \phi(p_T^{\text{MET}}, p_T^{i})) > 0.5$ for upto four leading jets with $p_T > 30$ GeV with $|\eta| < 4.7$. 
Motivation: Looking at VBF Higgs through a CNN

Invisible Higgs search at LHC

**Data-representation: high-level and low-level features**

Preprocessing

Network Performance

Result: Bounds on invisible branching ratio of Higgs

Back-up
Low-level: Tower-image

Pixel wise calorimeter energy deposits ($E_T$) converted into pictorial description like ‘tower-images’ as input to Convolutional Neural Networks.
Different resolution of calorimeter towers in central and forward regions
Low-level: Tower-image

- **Bin-size**: High-resolution (HR) $0.08 \times 0.08$ and a low-resolution (LR): $0.17 \times 0.17$, $\eta \in (-5, 5)$ and $\phi \in (-\pi, \pi)$
Low-level: Tower-image

- **Bin-size**: High-resolution (HR) $0.08 \times 0.08$ and a low-resolution (LR): $0.17 \times 0.17$, $\eta \in (-5, 5)$ and $\phi \in (-\pi, \pi)$

- Periodic in $\phi$
Low-level: Tower-image

- **Bin-size**: High-resolution (HR) $0.08 \times 0.08$ and a low-resolution (LR): $0.17 \times 0.17$, $\eta \in (-5, 5)$ and $\phi \in (-\pi, \pi)$

- **Padding**: padded at each $\phi$-boundary with rows from the opposite boundary.
Low-level: Tower-image

- **Bin-size**: High-resolution (HR) $0.08 \times 0.08$ and a low-resolution (LR): $0.17 \times 0.17$, $\eta \in (-5, 5)$ and $\phi \in (-\pi, \pi)$

- **Padding**: padded at each $\phi$-boundary with rows from the opposite boundary.

- **Size**: LR: $59 \times 45$, and HR: $125 \times 95$. 
High-level features: Event kinematics and QCD radiation

- **Kinematic**: Information about the event-kinematics from reconstructed objects

\[ \mathcal{K} \equiv (|\Delta \eta_{jj}|, |\Delta \phi_{jj}|, m_{jj}, \text{MET}, \phi_{\text{MET}}, \Delta \phi_{j_1}^{\text{MET}}, \Delta \phi_{j_2}^{\text{MET}}, \Delta \phi_{j_1+j_2}^{\text{MET}}) \]

- **Radiative**: Contains information about the QCD radiation pattern.

\[ \mathcal{R} \equiv (H_{T_{\eta_C}}^{mc} | \eta_C \in \mathcal{E}) \quad , \quad H_{T}^{mc} = \sum_{\eta<|\eta_C|} E_T \]

\( \mathcal{E} \): set of chosen \( \eta_C \)'s.

Vary \( \eta_C \) uniformly in the interval \([1,5]\) to get 16 \( H_{T}^{mc} \) variables.

- **Combined high-level feature space**: \( \mathcal{H} \)
Motivation: Looking at VBF Higgs through a CNN

Invisible Higgs search at LHC

Data-representation: high-level and low-level features

Preprocessing

Network Performance

Result: Bounds on invisible branching ratio of Higgs

Back-up
Low-level: Event-preprocessing

- Rotate along z-axis such that $\phi_0 = 0$.
  Two instances of $\phi_0 \in \{\phi_{MET}, \phi_{j1}\}$.

- Reflect along the $xy$-plane, such that the leading jet’s $\eta$ is always positive.

- After binning ($E_T$) and padding in LR and HR: $P_{MET}^{LR}$, $P_{MET}^{HR}$, $P_{J}^{LR}$ and $P_{J}^{HR}$
Low-level: Event-preprocessing

Averaged Images
Motivation: Looking at VBF Higgs through a CNN

Invisible Higgs search at LHC

Data-representation: high-level and low-level features

Preprocessing

**Network Performance**

Result: Bounds on invisible branching ratio of Higgs

Back-up
Receiver Operator Characteristics (ROC)

Quantification of classification power: ROC ⇒ Area Under Curve (AUC)
Receiver Operator Characteristics (ROC)

Quantification of classification power: ROC ⇒ Area Under Curve (AUC)

Low-level: $\mathcal{P}^{LR}_{\text{MET}}, \mathcal{P}^{HR}_{\text{MET}}, \mathcal{P}^{LR}_J$ and $\mathcal{P}^{HR}_J$ ⇒ CNNs

High-level: $\mathcal{K}$ (kinematic), $\mathcal{R}$ (QCD-radiative) and $\mathcal{H}$ (combination of the two previous spaces) ⇒ densely connected ANNs
Network Performance

**ROC: Low-level**

- $\mathcal{B}^{\text{R}}$: CNN
- $\mathcal{B}^{\text{R}}$: CNN
- $\mathcal{B}^{\text{R}}$: CNN
- $\mathcal{B}^{\text{R}}$: CNN
- $\mathcal{B}^{\text{R}}$: CNN

**ROC: High-level**

- $\mathcal{B}^{\text{R}}$: CNN
- $\mathcal{B}^{\text{R}}$: CNN
- $\mathcal{B}^{\text{R}}$: CNN
- $\mathcal{B}^{\text{R}}$: CNN
- $\mathcal{B}^{\text{R}}$: CNN

**AUC: Low-level**

- $\mathcal{B}^{\text{R}}$: CNN: 0.7576, 0.7779
- $\mathcal{B}^{\text{R}}$: CNN: 0.7576, 0.7732
- $\mathcal{B}^{\text{R}}$: CNN: 0.7663, 0.7864
- $\mathcal{B}^{\text{R}}$: CNN: 0.7661, 0.7826

**AUC: High-level**

- $\mathcal{B}^{\text{R}}$: CNN: 0.7042, 0.7007
- $\mathcal{B}^{\text{R}}$: CNN: 0.7177, 0.7095
- $\mathcal{B}^{\text{R}}$: CNN: 0.7359, 0.7319
- $\mathcal{B}^{\text{R}}$: CNN: 0.7663, 0.7864
Network Performance: Channel-wise outputs

- Harder to distinguish $S_{QCD}$ from the QCD dominated ($\sim 95\%$) background class (significant $S_{QCD}$ contamination in traditional analysis too)

- For the CNN, $W_{QCD}$ dominates over $Z_{QCD}$ in the first bin??
Network Performance: Channel-wise outputs

- Harder to distinguish $S_{QCD}$ from the QCD dominated (~95%) background class (significant $S_{QCD}$ contamination in traditional analysis too)

- For the CNN, $W_{QCD}$ dominates over $Z_{QCD}$ in the first bin??
  ⇒ Presence of calorimeter deposits of lepton in regions $|\eta| > 2.5$ or in the central regions when it is misidentified (including $\tau^{\pm}$).
Motivation: Looking at VBF Higgs through a CNN

Invisible Higgs search at LHC

Data-representation: high-level and low-level features

Preprocessing

Network Performance

Result: Bounds on invisible branching ratio of Higgs

Back-up
Bounds on $\text{B.R}(h^0 \rightarrow \text{inv})$

Reproduced CMS result at 36 fb$^{-1}$ (actual: BR < 0.25)

Expected 95% C.L median upper limit on the invisible branching ratio of SM Higgs with one and two sigma sidebands.
Bounds on B.R($h^0 \to \text{inv}$)

<table>
<thead>
<tr>
<th>Sl.No</th>
<th>Name</th>
<th>Description</th>
<th>Expected median upper-limit on B.R($h^0 \to \text{inv}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>L = 36 fb⁻¹</td>
</tr>
<tr>
<td>1.</td>
<td>$m_{ij}(\text{MET} &gt; 250 \text{ GeV})$</td>
<td>reproduced CMS shape analysis</td>
<td>0.226±0.093⁻0.063</td>
</tr>
<tr>
<td>2.</td>
<td>$</td>
<td>\Delta \eta_{ij}</td>
<td>(\text{MET} &gt; 250 \text{ GeV})$</td>
</tr>
<tr>
<td>3.</td>
<td>$m_{ij}(\text{MET} &gt; 200 \text{ GeV})$</td>
<td>$m_{ij}$ shape analysis with weaker cut</td>
<td>0.191±0.075⁻0.053</td>
</tr>
<tr>
<td>4.</td>
<td>$</td>
<td>\Delta \eta_{ij}</td>
<td>(\text{MET} &gt; 200 \text{ GeV})$</td>
</tr>
<tr>
<td>5.</td>
<td>$\mathcal{P}^{LR}_{j}-\text{CNN}$</td>
<td>Low-Resolution, $\phi_0 = \phi_{j_1}$</td>
<td>0.078±0.030⁻0.022</td>
</tr>
<tr>
<td>6.</td>
<td>$\mathcal{P}^{HR}_{j}-\text{CNN}$</td>
<td>High-Resolution, $\phi_0 = \phi_{j_1}$</td>
<td>0.070±0.027⁻0.020</td>
</tr>
<tr>
<td>7.</td>
<td>$\mathcal{P}^{LR}_{\text{MET}}-\text{CNN}$</td>
<td>Low-Resolution, $\phi_0 = \phi_{\text{MET}}$</td>
<td>0.092±0.037⁻0.025</td>
</tr>
<tr>
<td>8.</td>
<td>$\mathcal{P}^{HR}_{\text{MET}}-\text{CNN}$</td>
<td>High-Resolution, $\phi_0 = \phi_{\text{MET}}$</td>
<td>0.086±0.035⁻0.024</td>
</tr>
<tr>
<td>9.</td>
<td>$\mathcal{K}-\text{ANN}$</td>
<td>8 kinematic-variables</td>
<td>0.101±0.052⁻0.022</td>
</tr>
<tr>
<td>10.</td>
<td>$\mathcal{R}-\text{ANN}$</td>
<td>16 radiative $H_T^{NC}$ variables</td>
<td>0.138±0.055⁻0.039</td>
</tr>
<tr>
<td>11.</td>
<td>$\mathcal{H}-\text{ANN}$</td>
<td>Combination of $\mathcal{K}$ and $\mathcal{R}$ variables</td>
<td>0.094±0.038⁻0.026</td>
</tr>
</tbody>
</table>

- factor of three improvement, utilising the same amount of data.
- It can constrain many different BSM models severely.
### Bounds on $\text{B.R}(h^0 \rightarrow \text{inv})$

<table>
<thead>
<tr>
<th>Sl.No</th>
<th>Name</th>
<th>Description</th>
<th>Expected median upper-limit on $\text{B.R}(h^0 \rightarrow \text{inv})$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$L = 36 \text{ fb}^{-1}$</td>
</tr>
<tr>
<td>1.</td>
<td>$m_{ij}(\text{MET} &gt; 250 \text{ GeV})$</td>
<td>reproduced CMS shape analysis</td>
<td>$0.226^{+0.093}_{-0.063}$</td>
</tr>
<tr>
<td>2.</td>
<td>$</td>
<td>\Delta\eta_{ij}</td>
<td>(\text{MET} &gt; 250 \text{ GeV})$</td>
</tr>
<tr>
<td>3.</td>
<td>$m_{ij}(\text{MET} &gt; 200 \text{ GeV})$</td>
<td>$m_{ij}$ shape analysis with weaker cut</td>
<td>$0.191^{+0.075}_{-0.053}$</td>
</tr>
<tr>
<td>4.</td>
<td>$</td>
<td>\Delta\eta_{ij}</td>
<td>(\text{MET} &gt; 200 \text{ GeV})$</td>
</tr>
<tr>
<td>5.</td>
<td>$\mathcal{P}_{jLR}^\text{CNN}$</td>
<td>Low-Resolution, $\phi_0 = \phi_{j_1}$</td>
<td>$0.078^{+0.030}_{-0.022}$</td>
</tr>
<tr>
<td>6.</td>
<td>$\mathcal{P}_{jHR}^\text{CNN}$</td>
<td>High-Resolution, $\phi_0 = \phi_{j_1}$</td>
<td>$0.070^{+0.027}_{-0.020}$</td>
</tr>
<tr>
<td>7.</td>
<td>$\mathcal{P}_{jLR MET}^\text{CNN}$</td>
<td>Low-Resolution, $\phi_0 = \phi_{\text{MET}}$</td>
<td>$0.092^{+0.037}_{-0.025}$</td>
</tr>
<tr>
<td>8.</td>
<td>$\mathcal{P}_{jHR MET}^\text{CNN}$</td>
<td>High-Resolution, $\phi_0 = \phi_{\text{MET}}$</td>
<td>$0.086^{+0.035}_{-0.024}$</td>
</tr>
<tr>
<td>9.</td>
<td>$\mathcal{K}^\text{-ANN}$</td>
<td>8 kinematic-variables</td>
<td>$0.101^{+0.052}_{-0.022}$</td>
</tr>
<tr>
<td>10.</td>
<td>$\mathcal{R}^\text{-ANN}$</td>
<td>16 radiative $H_C^\text{TC}$ variables</td>
<td>$0.138^{+0.055}_{-0.039}$</td>
</tr>
<tr>
<td>11.</td>
<td>$\mathcal{H}^\text{-ANN}$</td>
<td>Combination of $\mathcal{K}$ and $\mathcal{R}$ variables</td>
<td>$0.094^{+0.038}_{-0.026}$</td>
</tr>
</tbody>
</table>

*▶ Pileup increases the upper-limit within $1\sigma$ errors for $\mathcal{P}_{jHR}^\text{CNN}$.***
Conclusion

- Posibility to replace decades old dependence on central-jet veto for the reduction of non-VBF backgrounds, in the meantime gaining significantly in performance.

- Low-level calorimeter image outperforms high-level physics motivated features.
  
  - High-level variables need reconstruction of events.  
    \(\implies\) Feasibility of CNN/ANN triggers for VBF?

- Minimally affected by pileup even without any mitigation.
Motivation: Looking at VBF Higgs through a CNN

Invisible Higgs search at LHC

Data-representation: high-level and low-level features

Preprocessing

Network Performance

Result: Bounds on invisible branching ratio of Higgs

Back-up
Event simulation details

- Modified version of Higgs Effective Field theory model
  ⇒ Higgs decays at parton level to two scalar dark matter particles for signal

- Finite top-mass: Reweight the Missing \( E_T(MET) \) distribution

- After preselection cuts: unweighted for Neural Network training

- Parton level cross-sections matched upto 4 and 2 jets for \( Z_{QCD} \) and \( W_{QCD} \), respectively
Details of data used in analysis

- Signal and background classes formed by mixing the channels with the expected proportions: \( k \times \sigma \times \epsilon_{\text{baseline}} \)

- **Shape-analysis** \((\text{MET} > 250 \text{ GeV})\):
  - Signal: 39% \( S_{\text{EW}} \) and the 61% \( S_{\text{QCD}} \)
  - Background: 54.43% \( Z_{\text{QCD}} \), 40.92% \( W_{\text{QCD}} \), 3.05% \( Z_{\text{EW}} \) and 1.58% \( W_{\text{EW}} \)
  - Expected number of background events at 36 fb\(^{-1}\) integrated luminosity, scaled for other luminosities.

- **Neural Network analysis** \((\text{MET} > 200 \text{ GeV})\):
  - Signal: 44.8% \( S_{\text{EW}} \) and the 55.2% \( S_{\text{QCD}} \)
  - Background: 51.221% \( Z_{\text{QCD}} \), 44.896% \( W_{\text{QCD}} \), 2.295% \( Z_{\text{EW}} \) and 1.587% \( W_{\text{EW}} \)
  - 100,000 training and 25,000 validation events for each class
  - Models completely agnostic to validation data
  - Further statistical analysis uses validation data scaled by different luminosities.

- Performed shape-analysis for \( \text{MET} > 200 \text{ GeV} \), for a better comparison.
High-level features: Kinematic

\( \text{MET} > 200 \text{ GeV} \)

\[ K \equiv \left( |\Delta \eta_{jj}|, \Delta \phi_{j1+2}^{\text{MET}}, \Delta \phi_{j1}^{\text{MET}}, \Delta \phi_{j2}^{\text{MET}}, \text{MET}, \phi_{\text{MET}}, \Delta \phi_{j1}^{\text{MET}}, \Delta \phi_{j2}^{\text{MET}} \right) \]

\( \text{MET} > 250 \text{ GeV} \)

\[ K \equiv \left( |\Delta \eta_{jj}|, m_{jj}, \text{MET}, \phi_{\text{MET}}, \Delta \phi_{j1}^{\text{MET}}, \Delta \phi_{j2}^{\text{MET}} \right) \]
High-level features: QCD-Radiative

\[ R \equiv (H_T^{\eta c} | \eta c \in \varepsilon) \]

\[ H_T^{\eta c} = \sum_{\eta < |\eta c|} E_T \]
After training for 20-1000 epochs, best performing network on the validation data chosen (for each of the 7 networks).

ANN architectures are inspired by the information bottleneck principle, closely related to coarse-graining in RG evolution.